



# Python을 이용한 냉동기 에너지소비량 예측 모델의 성능 개선 및 비교 평가

이철원<sup>1</sup> · 성남철<sup>2</sup> · 최원창<sup>3\*</sup>

<sup>1</sup>가천대학교 건축공학과 연구원, <sup>2</sup>가천대학교 친환경시스템연구센터 연구교수, <sup>3</sup>가천대학교 건축공학과 교수

## Performance Improvement and Comparative Evaluation of the Chiller Energy Consumption Forecasting Model Using Python

Lee, Cheol-Won<sup>1</sup> · Seong, Nam-Chul<sup>2</sup> · Choi, Won-Chang<sup>3\*</sup>

<sup>1</sup>Researcher, Department of Architect Engineering, Gachon University, Seongnam, Korea

<sup>2</sup>Research Professor, Eco-system Research Center, Gachon University, Seongnam, Korea

<sup>3</sup>Professor, Department of Architect Engineering, Gachon University, Seongnam, Korea

\*Corresponding author: Choi, Won-Chang, Tel: +82-31-750-5335, E-mail: wchoi@gachon.ac.kr

### ABSTRACT

In this study, Python is used to predict chiller energy consumption and improve the performance of forecasting models. The forecasting model used a random forest model and an artificial neural network model. To improve the performance of the forecasting model, the accuracy was evaluated by adjusting the number of inputs and the training data size. As a result, for the random forest model, the prediction performance allowed by the criteria was shown from the number of input variables to seven, and the CvRMSE improved the prediction performance by up to 23.91% by increasing the number of inputs. The training data size was shown to have acceptable predictive performance for the criterion at 80% and increased the training data size, improving the predictive performance by up to 14.08%. For artificial neural network (ANN) models, the predictive performance allowed by the criterion was shown to have a predictive performance with four inputs, and the CvRMSE improved by up to 14.90% by increasing the number of inputs. The training data size was shown to have acceptable predictive performance for the criterion at 70% and the maximum increase in the training data size resulted in improved predictive performance by up to 11.99% for CvRMSE. Comparing the two models, the artificial neural network model has better predictive performance than the random forest model, and the model for improving predictive performance is also more advantageous for the use of input variables and the adjustment of training data size.

**주요어** : 파이썬, 랜덤포레스트, 인공신경망, 예측 모델, 냉동기 에너지소비량

**Keywords**: Python, Random Forest, Artificial Neural Network, Forecasting Model, Chiller Energy Consumption

### OPEN ACCESS

Journal of KIAEBS 2021 June, 15(3): 252-264  
<https://doi.org/10.22696/jkiaeb.20210022>

pISSN : 1976-6483  
eISSN : 2586-0666

Received: April 19, 2021

Revised: June 11, 2021

Accepted: June 16, 2021

© 2021 Korean Institute of Architectural Sustainable Environment and Building Systems.



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 서론

건물에서 소비되는 에너지는 국가에서 사용하는 에너지의 40% 이상을 차지하며, 국내 건물의 전력 에너지소비량은 2013년에서 2019년 사이에 37% 증가하여 건물 에너지 소비의 절약을 위한 노력이 필요하다. 건물 에너지는 설계단계에서 에너지 절약적인 설계뿐만이 아닌 설비시스템의 적절한 운영, 제어 그리고 관리방안을 통하여 절약할 수 있다. 최근에는 BEMS (Building Energy Management System)와 같은 시스템의 도입이 활성화되면서 건물 에너지 사용량에 대한 다양한 정보를 제공하여 건물의 에너지를 효과적으로 관리가 가능해졌다. 건물 에너지 관리 방법 중에서 건축물의 에너지소비량과 수요의 정확한 예측은 설계단계에서부터 운영단계까지 에너지 성능 최적화를 위해서 반드시 필요하다. 설계단계에서는 시스템의 설계, 적정 설비 용량의 선정을 할 수 있으며, 운영단계에서는 최적제어, 적절한 운전 계획 등을 수립함으로써 에너지 성능을 향상시킬 수 있다.

이를 위해 최근에는 기계학습 방법을 이용하여 에너지 소비패턴을 분석하거나(Yoon et al., 2017) 에너지 소비 예측(Jeong and Chae, 2017)에 관한 연구들이 수행되고 있는데 대부분 건물 전체의 에너지소비량이나 건물 부하(Kim et al., 2019a)를 대상으로 에너지 예측을 수행하였다. 건물에서는 냉난방과 관련된 건축설비 시스템에서의 에너지의 절감 잠재력이 가장 크기 때문에 이들의 개별적인 에너지 소비 패턴을 이해하고 분석하여 정확한 예측이 요구되고 있으며 단일 기계학습 방법을 이용하여 예측하는 방법(Jeon and Kim, 2017; Kim et al., 2019b)이 주로 이루어지고 있어 여러 모델을 이용한 비교도 필요하다.

따라서 본 연구에서는 건물 에너지의 효과적인 관리를 위해 기계학습을 이용하여 에너지소비량을 정확하게 예측하는 연구를 수행하고자 한다. 예측은 건물에서 여름철의 냉방 에너지 소비 대부분을 차지하는 냉동기 에너지소비량을 대상으로 한다. Python을 기반으로 하여 기계학습 모델 중에서 랜덤포레스트 모델과 인공신경망 모델을 개발한 후 모델의 입력 조건들을 변화시켜가며 예측정확도를 산출하여 그 결과를 평가해가면서 예측 모델의 성능을 개선하였다. 또한 입력 조건들을 변화에 따른 두 모델의 비교를 통해 예측성능을 비교 분석하였다.

## 냉동기 데이터 생성 및 입력변수 선정

### 냉동기 운전 데이터 생성

본 연구에서 사용된 냉동기 운전 데이터는 선행된 연구(Seong et al., 2020)와 같이 업무용 표준건물의 시뮬레이션을 통해 생성된 데이터를 연구에 활용하였다. 업무용 표준건물은 DOE의 Building Energy Codes Program에서 제시한 표준건물 중 Commercial Prototype Building Models의 Large Office Building (Field et al., 2010; Deru et al., 2011)을 이용하였다. 표준건물을 시뮬레이션하는 과정에서 건물과 설비시스템의 주요 운전 스케줄, 내부 발열 조건, 냉방 설정온도는 선행연구(Seong et al., 2019)에서 활용한 것과 같이 국내의 일반적인 기준을 참고하여 Table 1과 같이 변경하였다. 기상 데이터는 서울지역의 기상자료인 TRY (Test Reference Year)format을 사용하였고 시뮬레이션 프로그램은 Energyplus 8.9를 사용

하였다. 데이터는 1시간 단위로 생성되도록 하였다. 총 데이터세트의 개수는 8760개이며 이중 냉방 기간인 5월부터 9월의 데이터 중 결측치인 0의 값을 제외한 3761개의 데이터세트를 연구에 사용하였다.

**Table 1.** Simulation Condition of Reference Building Large Scale Office (Seong et al., 2019)

Component	Features
HVAC Sizing	Auto Calculated
HVAC Operation Schedule	7:00~18:00
Internal Gain	Lighting 10.76 [W/m <sup>2</sup> ]
	People 18.58 [m <sup>2</sup> /person]
	Plug and Process 10.76 [W/m <sup>2</sup> ]
SetPoint	Cooling 26[°C]
Weather Data and Site Location	TRY Seoul (latitude: 37.57°N, longitude: 126.97°E)
Output Timestep	1 Hour

### 스피어만 등위 상관계수를 이용한 입력변수 선정

시뮬레이션 모델을 이용하여 생성된 다수의 데이터 중 기계학습 기반의 예측 모델의 입력 변수로 사용할 변수들을 선정하였다. 기계학습 기반의 예측 모델은 입력변수들의 중요도의 차이와 입력변수들의 개수에 따라 예측성능에 영향을 미친다(Shin and Park, 2017; Kim et al., 2019b). 따라서 본 연구에서는 예측 모델에 사용되는 입력변수들의 설정에 따른 예측성능의 차이를 확인하고, 변수들과 냉동기 에너지소비량 사이의 상관관계를 확인하기 위하여 스피어만 등위 상관계수(Spearman rank-order correlation coefficient)를 산출한 후 입력변수로 활용하였다.

Table 2는 생성된 입력변수들과 기계학습모델이 예측하여 출력하고자 하는 냉동기 에너지소비량 간의 스피어만 등위 상관계수 산출 결과를 나타낸다. 스피어만 등위 상관계수는 -1.0에서 +1.0 사이의 값을 가지며 양의 상관관계일 경우 +1에, 음의 상관관계일 경우 -1에 가까운 값을 가진다.

**Table 2.** Spearman rank-order correlation coefficient

Input value	Correlation Coefficient	Rank
Chilled water supply mass flow rate (kg/s)	0.988	1
Condenser water temperature (°C)	0.898	2
Outdoor dry-bulb temperature (°C)	0.893	3
Outdoor wet-bulb temperature (°C)	0.880	4
Outdoor dew point temperature (°C)	0.832	5
Outdoor relative humidity (%)	0.164	7
Operation hour	0.059	8
Operation state	0.039	9
Chilled water supply temperature (°C)	-0.672	6

냉동기 운전 여부를 결정하는 외기 조건 중에서 외기의 건구온도, 외기의 습구온도, 외기의 상대습도는 높은 상관관계를 보였으며, 냉동기 운전 값의 경우 냉수 공급 유량과 냉수 공급온도가 높은 상관관계를 보였다. 냉동기의 냉수 공급온도는 6.7°C의 고정값으로 설정하고 부하 변동에 따라 유량이 제어되도록 데이터를 생성하여 낮은 상관계수를 보였으며 냉수 공급 유량은 가장 높은 상관관계를 보였다.

입력변수는 양의 상관계수를 가지는 냉수 공급 유량(kg/s), 냉각수 온도(°C), 외기 건구온도(°C), 외기 습구온도(°C), 외기 노점온도(°C), 외기 상대습도(%), 가동시간, 가동날짜를 사용하였다.

## 예측 모델의 성능향상 및 평가 방법

본 연구에서는 Python을 이용하여 랜덤포레스트(Random Forest)와 인공신경망(Artificial Neural Network) 예측 모델을 개발하고 예측성능을 개선한 후 동일 조건에서 각 모델의 성능을 비교 분석하였다. Matlab과 Python 각 언어의 프로그래밍 특징들을 비교 분석한 연구(Ozgun et al., 2017)에서 Python이 Matlab보다 프로그래밍 작업이 더 간결하고 가독성이 좋기 때문에 학습 시에 버그가 감소하고 디버깅이 빠르며, The object-oriented programming (OOP)이 복잡한 Matlab 프로그램과는 달리 단순하고 유연한 내부 구조로 프로그래밍 과정의 접근성이 우수하다. 또한 Python은 오픈소스(open source) 개발 언어로 비용이 들지 않고 다른 프로그래밍 언어로 제작된 모듈들도 포함해 사용이 가능하다. 또한 코딩 시 변수를 사용할 때 변수 타입을 별도로 지정하지 않고 실행할 때 변수의 타입을 자동으로 검사를 실시한다. 컴파일하지 않고 인터프리터가 코드를 직접 한 줄씩 실행하는 방식으로 컴파일 과정이 필요 없고 실행 결과를 바로 확인하면서 코딩할 수 있다는 장점이 있다(Oliphant, 2007; Millman and Aivazis, 2011).

### 랜덤포레스트 모델

랜덤포레스트 예측 모델은 Python의 sklearn에서 제공하는 RandomForestRegressor를 사용하였으며 데이터의 계산, 그래프 작성, 통계 처리를 위한 모듈 설정을 실시하였다.

랜덤포레스트 모델의 예측성능 향상을 위해 입력변수의 개수와 훈련 데이터 크기를 변화시켜가며 예측 모델의 성능을 최적화하는 과정을 수행하였다. 입력변수의 개수는 상관관계가 가장 높은 냉수 공급 유량을 입력하는 것을 시작으로 차 순위의 변수들을 하나씩 순차적으로 추가하였으며 1개에서 8개까지 입력변수의 개수를 증가시켜가며 예측 결과를 비교하였다. 학습 데이터 크기는 50%에서 90%까지 10%씩 증가시켜가며 예측 결과를 비교하였다. 입력변수의 개수 변화 시에는 학습데이터의 크기는 80%로 고정하였고 학습데이터의 크기의 변화 시에는 입력변수의 개수를 8개로 고정하였다. Table 3은 랜덤포레스트 모델의 입력 파라미터 설정 조건을 나타낸다.

**Table 3.** Condition for Setting Input Parameters for Random Forest Model

Number of Input	1~8
Training Data Size	50~90%
Predict Target	Chiller Energy Consumption (kWh)

### 인공신경망 모델

인공신경망 예측 모델은 Python의 Tensorflow에서 제공하는 keras를 사용하였으며 데이터의 계산, 그래프 작성 그리고 통계처리를 위한 모듈 설정을 실시하였다. 인공신경망 모델의 예측성능 향상을 위해 랜덤포레스트와 동일한 방법으로 입력변수의 개수 증가와 학습데이터의 크기를 증가시켜가며 예측 모델의 성능을 최적화하는 과정을 수행하였다. 인공신경망의 성능을 최적화하는 과정에서는 뉴런(neuron)의 수도 함께 고려하여야 하는데 입력변수의 개수와 학습데이터의 크기를 변화할 때의 뉴런의 수는 60개로 고정하였다. 뉴런의 수를 변경하여가며 모델의 성능을 최적화할 때는 뉴런의 수를 10개부터 100개까지 10개씩 증가시켰다. 뉴런의 수를 변화시킬 때의 입력변수 개수는 8개, 훈련 데이터의 크기는 80%로 고정하였다. 인공신경망 모델의 구조는 앞서 수행된 연구(Kim et al., 2019b)를 참고하여 입력층, 은닉층, 출력층으로 구성되며 Learning Rate 0.001로, 학습 횟수(Epoch)는 100회로 설정하였으며 변경 없이 모든 경우에 일괄 적용하였다.

Table 4는 인공신경망 모델의 입력 파라미터 설정 조건을 나타낸다.

**Table 4.** Condition for Setting Input Parameters for Artificial Neural Network Model

Number of Input	1~8
Training Data Size	50~90%
Number of Neuron	10~100
Predict Target	Chiller Energy Consumption (kWh)

### 예측 모델 평가 방법

ASHRAE's Guideline 14 (ASHRAE, 2002)에 의하면 계산 결과의 경우 데이터 비교 결과가 Table 5의 범위 안에 해당하면 적절하다고 판단하여 신뢰할 수 있는 결과라고 보고 있다. 본 연구에서는 CvRMSE (Coefficient of Variation of Root Mean Square Error)를 사용하였다. 각 모델의 입력 조건을 변화시킴에 따라서 10회씩 예측한 후 CvRMSE의 최대, 최소, 평균을 구하여 최종적으로 모델의 예측성능을 평가하였다.

**Table 5.** Acceptable Calibration Tolerances (ASHRAE, 2002)

Calibration Type	Index	Acceptable Value*
Monthly	MBEmonth	± 5%
	Cv (RMSEmonth)	
Hourly	MBEhourly	±10%
	Cv (RMSEhourly)	

\*Lower values indicate better calibration

## 예측 결과 및 분석

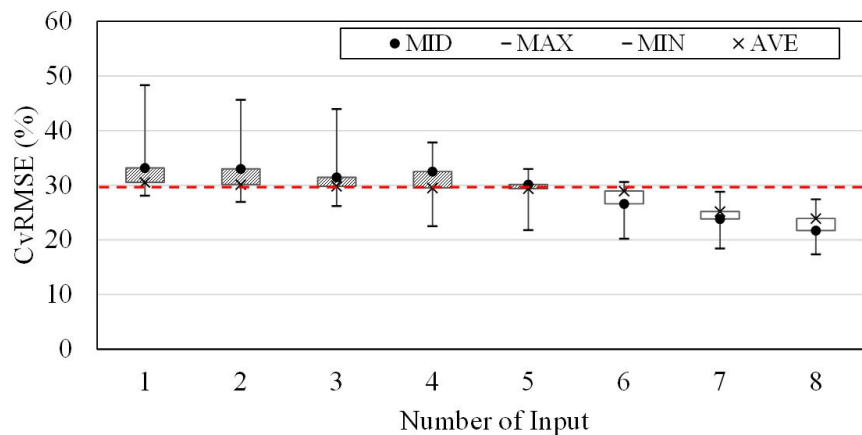
### 랜덤포레스트 모델의 예측성능 결과

Table 6은 랜덤포레스트 모델의 입력개수 변화에 따른 CvRMSE를 나타낸다. 입력개수를 3개로 설정하였을 때부터 CvRMSE의 평균이 29.80%로 기준인 30%를 초과하지 않지만 최댓값이 43.94%이므로 기준에 만족하는 예측성능을 확보할 수 있다고 보기는 힘들다고 판단된다. 최솟값, 최댓값, 평균값이 모두 기준을 만족하는 경우는 입력개수가 7개 이상으로 설정하였을 때부터로 본 연구에서 랜덤포레스트모델을 이용하여 냉동기 에너지소비량을 예측하기 위해서는 입력개수를 7개 이상으로 구성한 모델이 기준 내에 허용되는 예측성능을 갖는 모델이라고 할 수 있다. 입력변수의 개수를 7개에서 8개로 증가시키면 모델의 평균 예측정확도를 1.31% 더 향상할 수 있었다.

Figure 1은 입력개수 변화에 따른 예측정확도를 도식화 한 것으로 입력개수를 증가시켜가며 예측정확도를 평가한 결과 비교적 상관관계가 낮은 입력변수라도 예측성능에 영향을 미치며, 예측정확도인 CvRMSE는 입력변수의 개수가 증가함에 따라 감소하여 예측성능이 향상되는 것으로 나타났다.

**Table 6.** Accuracy of Random forest models according to change number of inputs

Number of Input	Maximum	Minimum	Average	Standard Deviation
1	48.31	28.12	30.53	2.87
2	45.62	26.98	30.11	2.96
3	43.94	26.21	29.80	1.72
4	37.83	22.51	29.50	3.32
5	32.98	21.82	29.38	0.85
6	30.62	20.23	28.93	2.47
7	28.82	18.46	25.22	1.30
8	27.43	17.39	23.91	2.22



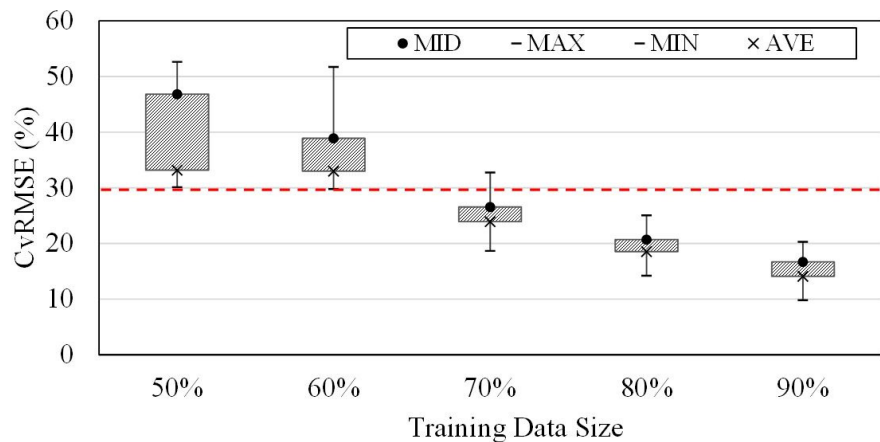
**Figure 1.** Accuracy of Random forest models according to change number of inputs

Table 7은 랜덤포레스트 모델의 훈련 데이터의 크기의 변화에 따른 예측 결과의 CvRMSE 값을 나타낸다. 훈련 데이터의 크기가 70%로 설정하였을 때 CvRMSE의 평균값이 23.91%로 기준을 만족하지만 최댓값이 32.78%로 기준을 초과하여 예측성능을 확보하였다고 보기는 어렵다. 최솟값, 최댓값, 평균값이 모두 기준을 만족하는 경우는 학습데이터의 크기를 80%로 조정된 모델로 본 연구에서 랜덤포레스트모델을 이용하여 냉동기 에너지소비량을 예측하기 위해서는 학습데이터의 크기를 80% 이상으로 구성한 모델이 기준 내에 허용되는 예측성능을 갖는 모델이라고 할 수 있다. 학습 데이터 크기를 80%에서 90%까지 증가시키면 모델의 평균 예측정확도는 4.48% 더 향상되었다.

Figure 2는 학습 데이터 크기에 따른 랜덤포레스트모델의 예측정확도를 도식화 한 것으로 학습 데이터 크기를 증가시키며 예측정확도를 평가한 결과 예측정확도인 CvRMSE는 훈련 데이터의 크기가 증가함에 따라 감소하여 예측성능이 향상되는 것으로 나타났다. 훈련 데이터의 크기가 증가할수록 많은 데이터를 훈련에 사용하여 더 높은 예측성능을 확보할 수 있지만, 훈련에 사용되는 데이터의 양이 증가하는 만큼 예측 모델의 예측 기간이 단축된다. 기계학습 기반의 예측 모델은 예측 기간을 고려하여 학습데이터와 평가데이터의 크기의 비율을 적절히 조정하는 것이 중요하다 판단된다.

**Table 7.** Accuracy of Random forest models according to change training data size

Training Data Size	Maximum	Minimum	Average	Standard Deviation
50	52.64	30.12	33.15	13.79
60	51.73	29.76	33.01	5.93
70	32.78	18.66	23.91	2.52
80	25.07	14.23	18.56	2.11
90	20.29	9.82	14.08	2.59



**Figure 2.** Accuracy of Random forest models according to change training data size

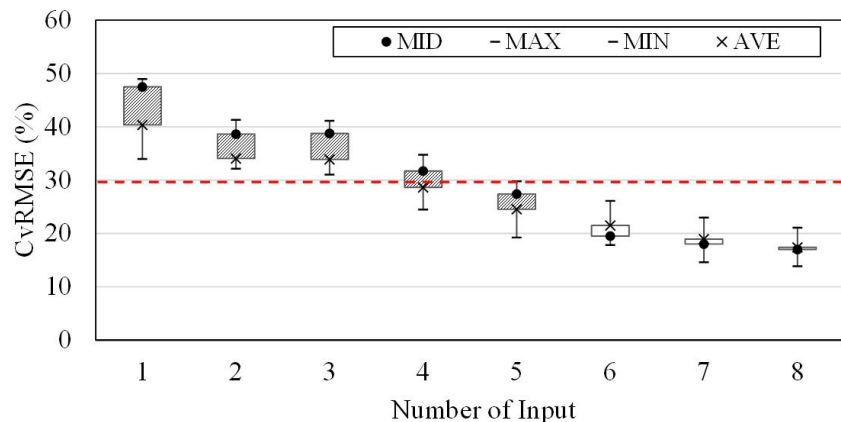
### 인공신경망 모델의 예측성능

Table 8은 인공신경망 모델의 입력변수의 개수 변화에 따른 예측 결과의 CvRMSE 값을 나타낸다. 인공신경망 모델도 랜덤포레스트모델과 마찬가지로 입력변수의 개수가 증가할수록 예측성능 또한 향상됨을 확인할 수 있었다. 인공신경망 모델은 입력변수를 3개로 구성한 모델부터 CvRMSE의 평균값은 26.90%로 예측성능 기준을 만족하지만 최댓값이 38.39%로 기준을 초과하여 예측성능을 확보하였다고 보기는 어렵다. 입력변수의 개수가 4개로 설정했을 때부터 최솟값, 최댓값, 평균값이 모두 기준을 만족하는 예측성능을 나타냈으며 랜덤포레스트모델에서 8개의 모든 입력변수의 개수를 사용한 모델보다 평균 예측정확도가 5.48% 향상되었다. 따라서 인공신경망 모델의 경우 같은 조건에서 랜덤포레스트모델 보다 적은 입력변수로 모델을 구성하여도 우수한 예측성능을 확보할 수 있다고 판단된다.

Figure 3은 입력개수 변화에 따른 예측정확도를 도식화 한 것으로 인공신경망 모델도 랜덤포레스트모델과 마찬가지로 비교적 상관관계가 낮은 입력변수라도 예측성능에 영향을 미치며, 예측정확도인 CvRMSE는 입력변수의 개수가 증가함에 따라 감소하여 예측성능이 향상되는 것으로 나타났다. 8개의 모든 입력변수를 사용하여 모델을 구성하였을 때 예측정확도는 4개의 입력변수로 모델을 구성하였을 때보다 3.58% 더 향상되었다.

**Table 8.** Accuracy of ANN models according to change number of inputs

Number of Input	Maximum	Minimum	Average	Standard Deviation
1	102.82	76.57	89.26	2.20
2	72.37	37.98	46.85	8.51
3	38.89	22.51	26.90	5.34
4	23.79	14.02	18.48	4.47
5	23.02	13.83	18.13	2.42
6	22.37	12.87	17.40	2.37
7	17.28	13.74	15.96	0.91
8	16.63	12.06	14.90	0.72



**Figure 3.** Accuracy of ANN models according to change number of inputs



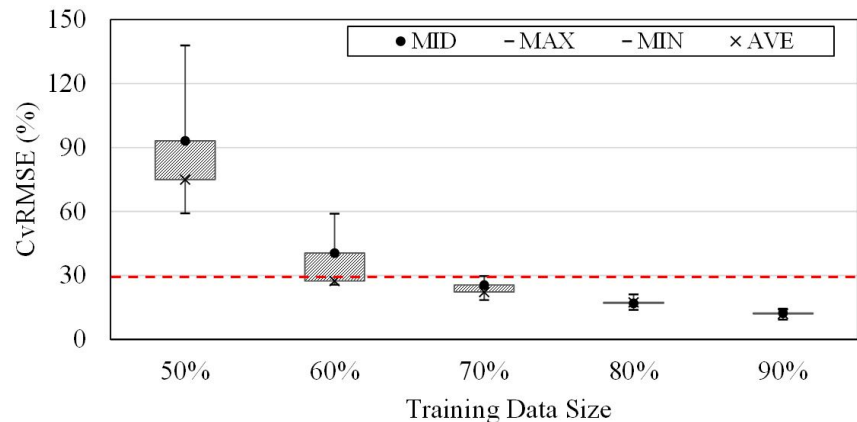
Table 9는 인공신경망 모델의 학습데이터의 크기의 변화에 따른 예측 결과의 CvRMSE 값을 나타낸다. 인공신경망 모델에서도 랜덤포레스트 모델과 마찬가지로 학습데이터의 크기가 증가함에 따라 예측성능도 향상되는 것을 확인할 수 있었다.

학습데이터의 크기가 60%로 설정하였을 때 CvRMSE의 평균값이 27.40%로 기준을 만족하지만 최댓값이 59.00%로 기준을 초과하여 예측성능을 확보하였다고 보기는 어렵다. 최솟값, 최댓값, 평균값이 모두 기준을 만족하는 경우는 학습데이터의 크기를 70%로 조정된 모델로 본 연구에서 인공신경망 모델을 이용하여 냉동기 에너지소비량을 예측하기 위해서는 학습데이터의 크기를 70% 이상으로 구성한 모델이 기준 내에 허용되는 예측성능을 갖는 모델이라고 할 수 있다. 학습데이터의 크기를 70%로 하였을 때보다 학습데이터의 크기를 80%로 증가시키면 모델의 평균 예측정확도는 4.9% 더 향상되었고, 학습데이터의 크기를 90%로 증가시키면 모델의 평균 예측정확도는 10.31% 더 향상되었다.

Figure 4는 학습데이터 변화에 따른 예측정확도를 도식화 한 것으로 학습 데이터 크기를 증가시켜가며 예측정확도를 평가한 결과 예측정확도인 CvRMSE는 랜덤포레스트 모델과 마찬가지로 훈련 데이터의 크기가 증가함에 따라 감소하여 예측성능이 향상되는 것으로 나타났다. 하지만 랜덤포레스트 모델과 비교하였을 때 학습데이터의 크기가 50%인 경우 오히려 예측성능은 더 좋지 않지만, 학습데이터의 크기를 60%로 하였을 때부터 예측성능이 급격히 향상되는 것을 확인할 수 있었다.

**Table 9.** Accuracy of ANN models according to change training data size

Training Data Size	Maximum	Minimum	Average	Standard Deviation
50	137.87	59.18	75.02	18.81
60	59.00	25.70	27.40	13.24
70	29.79	18.53	22.30	3.20
80	21.12	13.87	17.40	0.42
90	14.43	9.4	11.99	0.31



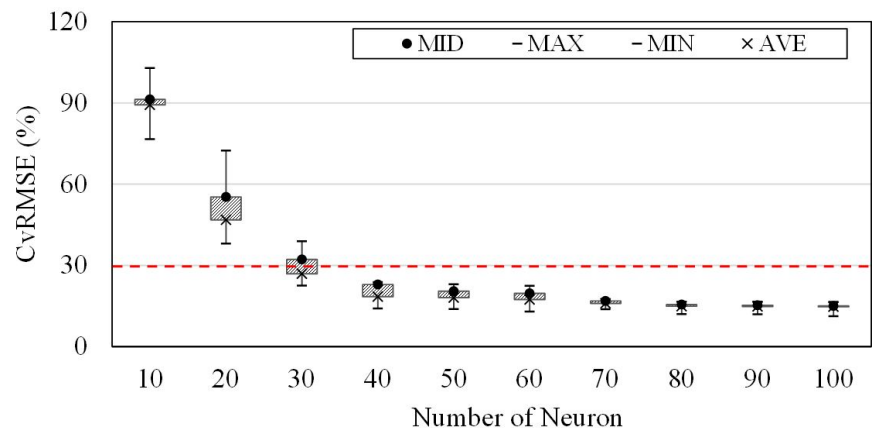
**Figure 4.** Accuracy of ANN models according to change training data size

인공신경망 모델은 추가로 뉴런 수 조정을 통해 모델의 예측성능을 향상할 수 있다. Table 10은 인공신경망 모델의 뉴런 개수의 변화에 따른 예측 결과의 CvRMSE 값을 나타낸다. 뉴런의 개수가 40개 이상일 경우, 모든 값이 기준에 만족하였다. 뉴런의 개수를 80개 이상에서는 뉴런의 수를 증가시켜도 CvRMSE값이 약 0.1% 정도만 증가하여 예측성능에 큰 차이를 보이지 않았다.

Figure 5는 뉴런 수 변화에 따른 예측정확도를 도식화 한 것으로 뉴런의 개수를 증가시키며 예측정확도를 평가한 결과 CvRMSE는 뉴런의 개수가 증가함에 따라 감소하여 예측성능이 향상되는 것으로 나타났다. 인공신경망 모델은 뉴런 수의 증가에 따라서 예측성능이 향상되지만, 일정 수준의 예측성능까지 도달한 후에는 뉴런의 수가 모델의 예측성능에 크게 영향을 주지 않는 것으로 판단되며 뉴런 수 조정에 의해 일정한 예측성능을 확보한 후에는 학습에 소요되는 시간을 감소시키는 것이 더 유리하다.

**Table 10.** Accuracy of ANN models according to change number of neuron

Number of Neuron	Maximum	Minimum	Average	Standard Deviation
10	102.82	76.57	89.26	2.20
20	72.37	37.98	46.85	8.51
30	38.89	22.51	26.90	5.34
40	23.79	14.02	18.48	4.47
50	23.02	13.83	18.13	2.42
60	22.37	12.87	17.40	2.37
70	17.28	13.74	15.96	0.91
80	16.63	12.06	14.90	0.72
90	16.58	11.93	14.86	0.44
100	16.53	11.23	14.78	0.23



**Figure 5.** Accuracy of ANN models according to change number of neuron

### 랜덤포레스트 모델과 인공신경망 모델 비교 분석

랜덤포레스트 모델과 인공신경망 모델의 예측성능을 비교하기 위하여 입력변수의 개수 변화에 따른 예측성능과 학습데이터의 변화에 따른 예측성능을 비교 평가하였다.

Table 11은 랜덤포레스트 모델과 인공신경망 모델의 입력변수의 개수 변화에 따른 CvRMSE의 평균값을 나타낸다. 랜덤포레스트 모델과 인공신경망 모델 모두 입력변수의 개수를 3개로 하였을 때부터 기준에 적합한 예측성능을 보였다. 그러나 입력변수의 개수를 3개에서 8개로 증가시키었을 때 인공신경망 모델이 랜덤포레스트 모델보다 9.01% 더 우수한 예측성능을 나타내었으며 랜덤포레스트 모델은 예측정확도가 5.89% 향상되는 반면 인공신경망 모델은 예측정확도가 12% 향상되어 입력변수 증가에 따른 예측성능의 개선 효과는 인공신경망 모델이 더 우수한 것으로 나타났다. 따라서 인공신경망 모델이 입력변수 증가에 따른 예측성능 개선 효과도 훨씬 더 크고, 적은 수의 입력변수로 모델을 구성하여도 우수한 예측성능을 확보할 수 있다.

**Table 11.** Accuracy comparison between random forest and artificial neural network model according to change number of inputs

	Number of Input							
	1	2	3	4	5	6	7	8
Random Forest	30.53	30.11	29.80	29.50	29.38	28.93	25.22	23.91
Artificial Neural Network	89.26	46.85	26.90	18.48	18.13	17.40	15.96	14.90

Table 12는 랜덤포레스트 모델과 인공신경망 모델의 학습데이터의 크기 변화에 따른 CvRMSE의 평균값들을 나타낸다. 랜덤포레스트 모델은 학습데이터의 크기를 70%로 조정하였을 때부터 기준에 만족하는 예측성능을 보여주지만, 인공신경망 모델은 학습데이터를 60%로 조정하였을 때부터 기준에 만족하는 예측성능을 보여준다. 학습데이터의 크기가 70%일 때 예측정확도는 인공신경망 모델이 1.60% 우수하였고 학습데이터의 크기가 80%일 때 예측정확도는 인공신경망 모델이 1.16% 우수하였으며 학습데이터의 크기가 90%일 때 예측정확도는 2.09% 더 우수하였다. 학습데이터의 크기를 70%에서 90%로 증가시켰을 때 인공신경망 모델이 랜덤포레스트 모델보다 2.09% 더 우수한 예측성능을 나타내었으며 랜덤포레스트 모델은 예측성능이 9.83% 향상됐지만 인공신경망 모델은 예측성능이 10.31% 향상되어 학습 데이터 크기 변화에 따른 예측성능 개선 효과도 훨씬 더 우수한 것으로 나타났다. 따라서 인공신경망 모델이 학습데이터 증가에 따른 예측성능 개선 효과도 훨씬 더 크고 학습데이터와 평가데이터의 비율의 크기 조정과 예측 기간 선정에서 훨씬 더 유리하다.

**Table 12.** Accuracy comparison between random forest and artificial neural network model according to change training data size

	Training Data Size				
	50%	60%	70%	80%	90%
Random Forest	33.15	33.01	23.91	18.56	14.08
Artificial Neural Network	75.02	27.40	22.30	17.40	11.99

## 결론

본 연구에서는 Python 프로그래밍을 기반으로 기계학습 모델들을 이용하여 냉동기의 에너지소비량을 정확하게 예측하기 위해서 입력변수의 개수 변화와 학습 데이터 크기 조정이 각 모델의 예측성능변화에 미치는 영향을 비교하고 평가하였다. 기계학습 모델은 랜덤포레스트 모델과 인공신경망 모델을 이용하였으며 결과들을 통해 도출된 결론은 다음과 같다.

랜덤포레스트 모델과 인공신경망 모델 모두 출력값과 비교적 상관관계가 낮은 입력변수일 지라도 입력변수의 개수를 증가시킴으로써 예측성능을 향상할 수 있었고, 학습데이터의 비율을 높일수록 예측성능을 향상할 수 있었다. 랜덤포레스트 모델의 경우 입력변수의 개수를 7개로 하였을 때부터 기준에 허용되는 예측성능을 갖는 것으로 나타났으며 입력변수의 개수를 최대로 증가시켜 CvRMSE는 23.91%까지 모델의 예측성능이 개선되었다. 학습데이터의 크기는 80%로 하였을 때 기준에 허용되는 예측성능을 갖는 것으로 나타났으며 학습데이터의 크기를 최대로 증가시켜 CvRMSE가 14.08%까지 모델의 예측성능이 개선되었다.

인공신경망 모델의 경우 입력변수의 개수를 4개로 하였을 때부터 기준에 허용되는 예측성능을 갖는 것으로 나타났으며 입력변수의 개수를 최고로 증가시켜 CvRMSE가 최대 14.90%까지 모델의 성능이 개선되었다. 학습데이터의 크기는 70%로 하였을 때 기준에 허용되는 예측성능을 갖는 것으로 나타났으며 학습데이터의 크기를 최대로 증가시켜 CvRMSE가 최대 11.99%까지 모델의 성능이 개선되었다. 인공신경망 모델은 뉴런 수의 증가에 따라서 예측성능이 향상되지만, 일정 수준의 예측성능까지 도달한 후에는 뉴런의 수가 모델의 예측성능에 크게 영향을 주지 않아 뉴런 수 조정에 의해 일정한 예측성능을 확보한 후에는 학습에 소요되는 시간을 감소시키는 것이 더 유리하다.

입력변수 증가에 따른 예측성능의 개선 효과와 학습데이터 증가에 따른 예측성능 개선 효과를 비교하여 분석한 결과 냉동기 에너지소비량 예측에는 랜덤포레스트 모델보다 인공신경망 모델이 더 우수한 예측성능을 갖는 것으로 나타났으며 모델을 입력변수의 사용과 학습데이터의 크기 조정에도 훨씬 더 유리하다.

후속 연구에서는 모델의 예측성능에 영향을 줄 수 있는 다른 파라미터의 조정을 통한 성능 평가가 필요하다. 또한 신경망 모델은 CNN (Convolutional Neural Network), RNN (Recurrent Neural Network), LSTM (Long Short-Term Memory model)과 같은 다양한 구조를 가진 신경망 모델이 존재하므로 이와 같은 모델을 활용한 연구도 수행되어야 할 것이다. 기계학습 기반의 에너지소비량 예측과 관리방안을 현장에 적용하기 위해서는 시뮬레이션 모델을 이용하여 생성된 데이터가 아닌 실측 데이터를 활용하여 모델의 예측성능 검증이 더 유용할 것으로 판단되며, 건물에서 에너지가 소비되는 공기조화기, 보일러, 펌프, 송풍기, 냉각탑과 같은 여러 설비시스템에 대하여 에너지소비량 예측 모델을 적용하고 예측성능을 평가해야 한다. 최종적으로는 개발된 모델을 기반으로 하는 운영과 제어에 관한 연구를 진행할 예정이다.

## 후기

이 논문은 가천대학교의 연구활성화 장학금 지원에 의한 연구임.

## References

1. Field, K., Deru, M., Studer, D. (2010). Using DOE commercial reference buildings for simulation studies. *Proceedings of SimBuild*, 4(1), 85-93.
2. Jeon, B.K., Kim, E.J. (2017). Short-Term Load Prediction Using Artificial Neural Network Models. *Korean Journal of Air-Conditioning and Refrigeration Engineering*, 29(10), 497-503.
3. Jeong, J.H., Chae, Y.T. (2017). Assessment of Input Variable Importance and Machine Learning Model Selection for Improving Short Term Load Forecasting on Different Building Types. *Journal of Korean Institute of Architectural Sustainable Environment and Building Systems*, 11(6), 586-598.
4. Kim, J.H., Seong, N.C., Choi, W. (2019a). Cooling load forecasting via predictive optimization of a nonlinear autoregressive exogenous (NARX) neural network model. *Sustainability*, 11(23), 6535.
5. Kim, J.H., Seong, N.C., Choi, W. (2019b). Modeling and optimizing a chiller system using a machine learning algorithm. *Energies*, 12(15), 2860.
6. Millman, K.J., Aivazis, M. (2011). Python for Scientists and Engineers. *Computing in Science & Engineering*, 13(2), 9-12.
7. Oliphant, T.E. (2007). Python for Scientific Computing. *Computing in Science & Engineering*, 9(3), 10-20.
8. Ozgur, C., Colliau, T., Rogers, G., Hughes, Z., Myer, T.B. (2017). MatLab vs. Python vs. R. *Journal of Data Science*, 15(3), 355-372.
9. Seong, N.C., Kim, J.H., Choi, W. (2019). Optimal Control Strategy for Variable Air Volume Air-Conditioning Systems Using Genetic Algorithms. *Sustainability*, 11(18), 5122.
10. Seong, N.C., Kim, J.H., Choi, W.C. (2020). Determination of optimal variables for chilled water loop in central air-conditioning system using genetic algorithms. *Journal of Korean Institute of Architectural Sustainable Environment and Building Systems*, 14(1), 66-79.
11. Shin, H.S., Park, C.S. (2017). Development of a Machine Learning Model for a Chiller using Random Forest Algorithm and Data Pre-processing. *Journal of the Architectural Institute of Korea Structure & Construction*, 33(9), 67-74.
12. Yoon, Y.R., Shin, S. H., Moon, H. J. (2017). Analysis of Building Energy Consumption Patterns according to Building Types Using Clustering Methods. *Journal of The Korean Society of Living Environmental System*, 24(2), 232-237.
13. Deru, M., Field, K., Studer, D., Benne, K., Griffith, B., Torcellini, B., Liu, B., Halverson, M., Winiarski, D., Rosenberg, M., Yazdani, M., Huang, J., Crawley, D. (2011). U.S. Department of Energy Commercial Reference Building Models of the National Building Stock (NREL/TP-5500-46861). National Renewable Energy Laboratory Technical Report.
14. ASHRAE. (2002). ASHRAE Guideline 14-2002: measurement of energy and demand savings. American Society of Heating, Refrigerating and Air-conditioning Engineers, Atlanta, GA.
15. Commercial Prototype Building Models-Building Energy Codes Program. Available at: [https://www.energycodes.gov/development/commercial/prototype\\_models](https://www.energycodes.gov/development/commercial/prototype_models).